



面向 21 世纪 课程教材
信息管理与信息系统专业教材系列

互联网信息资源检索与利用

第 3 版

符绍宏 雷菊霞 邓瑞丰 高冉 © 编著



清华大学出版社



面向 21 世纪课程教材
信息管理与信息系统专业教材系列

互联网信息资源检索与利用

第 3 版

符绍宏 雷菊霞 邓瑞丰 高冉 编著



清华大学出版社
北京

内 容 简 介

互联网信息的无限、无序、优劣混杂、缺乏统一的组织与控制等特点,给人们查找和利用信息造成了一定的不便。在互联网这个浩瀚、动荡的信息海洋中,准确、及时、有效地查找并获取所需要的信息,对所有网络用户都是十分重要的,同时也非常具有挑战性。本书面向互联网用户,针对检索与获取对用户研究开发、学习有价值的网络信息的需求,系统地介绍互联网信息资源的特点、种类、信息利用价值及评价、鉴定标准;介绍互联网信息检索的基础知识,各类型常用网络信息检索工具的功能、特点、种类、信息利用价值及使用和检索方法;详细介绍学术研究性信息、公共信息及各类专门信息的网上检索平台及查找方法。本书对培养网络信息查找和利用能力,提高在网络环境下生存和发展水平具有使用价值。

在上一版的基础上,本书结合网络检索工具的发展变化情况对相关内容进行了更新,可以作为高等院校网络信息检索课程的教材,也可以作为广大网络用户上网查找信息的实用参考书和指南。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

互联网信息资源检索与利用/符绍宏等编著.--3版.--北京:清华大学出版社,2012.9

(面向21世纪课程教材·信息管理 with 信息系统专业教材系列)

ISBN 978-7-302-29224-1

I. ①互… II. ①符… III. ①互联网络—情报检索—高等学校—教材 IV. ①G354.4

中国版本图书馆CIP数据核字(2012)第143360号

责任编辑:高晓蔚

封面设计:傅瑞学

责任校对:王荣静

责任印制:李红英

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦A座 邮 编:100084

社总机:010-62770175 邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

印 装 者:北京国马印刷厂

经 销:全国新华书店

开 本:185mm×230mm 印 张:21 插 页:1 字 数:448千字

版 次:2000年6月第1版 2012年9月第3版 印 次:2012年9月第1次印刷

印 数:1~5000

定 价:35.00元

产品编号:041665-01

前言 PREFACE

目前,互联网已经发展成为人类信息环境的重要组成部分,是人类交流信息、获取信息的重要渠道。网络信息检索也已经成为人们信息生活中非常重要的内容。有关研究调查显示,查阅资料、浏览新闻和即时通讯是网民们最热衷的上网行为。然而互联网信息资源的无限、无序、优劣混杂、缺乏统一的组织与控制等特点给人们查找和利用信息造成了很大的不便。人们在面对互联网这个海量信息世界时都会感到由信息过载(information overloaded)引发的困惑和茫然,要在这个浩瀚、动荡的信息海洋中,准确、及时、有效地查找并获取与自身信息需求相关、切合需要和适用的信息,对所有网络用户都是十分重要的,同时也非常具有挑战性。

IREsearch《2010—2011 年中国搜索引擎用户行为研究报告》说明,搜索已经成为中国用户最常使用的网络应用服务之一,近八成(76.9%)的网络用户接触搜索引擎的时间在3年以上,其中使用搜索引擎时间在3~5年、5~8年、8年以上的用户比例分别为21.5%、29.3%、26.1%。网络用户使用搜索引擎的时间在增长,表明搜索引擎在人们日常生活中的地位日益重要,搜索引擎的强大功能和效用正在被广泛认知并挖掘。但相对用户需求而言,不同年龄段的用户差异明显,亟待开发适应个性化搜索需求的定制化、个性化搜索引擎技术和搜索工具,网络搜索业所蕴藏的巨大财富还有待大力挖掘。同时,搜索引擎面临着其他互联网产品的激烈竞争,也需要搜索技术的变革来进一步拓展市场规模。

本书面向互联网用户在研究开发、经营管理、工作和学习等实际生活中产生的各种信息需求,系统介绍互联网信息资源的特点、种类、信息利用价值及评价鉴选标准,互联网信息检索的基础知识、各类型常用网络检索工具及各类学术信息,专门信息网上检索系统等的功能、特点和查找使用方法。对培养网络信息查找和利用能力,提高在网络环境下的生存和发展水平具有使用价值。

本书从2000年开始出版第1版,应市场需求,于2005年出版第2版。本书历经近20次印刷,从很多读者的来信中我们了解到,本书受到了广大读者的肯定,得到了各高等院校教学部门和企业用户的认可与广泛使用。在此,我们特向广大读者表示衷心的感谢。

本次修订在第2版的基础上结合网络环境、网络资源以及检索工具的发展变化情况对相关内容进行了更新。在修订和编写的过程中,北京师范大学管理学院在读研究生武莹同学也参与了部分工作,在此表示感谢。

限于编者学术水平和视野有限,加上当今的网络环境下的信息资源、检索平台又经常发生更新和变化,所以书中还会有不少疏漏之处,欢迎业界同行批评、指正,也恳请各位读者多多包涵。

符绍宏

2012年5月于北京师范大学

目 录 CONTENTS

第 1 章 互联网信息资源	1
1.1 互联网上的信息资源	1
1.1.1 网络上的信息内容	1
1.1.2 网上不容易获取的信息内容	2
1.2 互联网信息资源的特点及局限性	3
1.2.1 互联网信息资源的特点	3
1.2.2 互联网信息资源的局限性	4
1.3 互联网上信息资源的种类	4
1.3.1 Web 信息资源	5
1.3.2 Telnet 资源	6
1.3.3 FTP 信息资源	7
1.3.4 用户组信息资源	8
1.4 互联网信息资源的评价	9
1.4.1 评价的意义	9
1.4.2 评价的标准	10
1.4.3 评价方法	12
思考题	13
第 2 章 互联网信息检索概论	14
2.1 互联网信息检索的类型和特点	14
2.1.1 互联网信息检索的类型	15
2.1.2 互联网信息检索的特点	20
2.2 互联网信息检索原理之一——信息组织	21
2.2.1 文档分析	22
2.2.2 建立索引	24
2.2.3 归类和聚类	26
2.2.4 元数据	27
2.3 互联网信息检索原理之二——信息查询	30

2.3.1 布尔检索模型	30
2.3.2 扩展布尔模型	31
2.3.3 向量空间模型	32
2.3.4 概率检索模型	36
2.4 互联网检索工具的性能评价	36
2.4.1 收录范围	36
2.4.2 检索功能	37
2.4.3 检索效率	37
2.4.4 检索结果的处理和展示	38
2.4.5 用户界面设计	39
思考题	39
第3章 互联网检索工具(一)——搜索引擎	40
3.1 搜索引擎概述	40
3.1.1 搜索引擎的发展简史	40
3.1.2 搜索引擎的工作原理	42
3.2 搜索引擎的类型	47
3.2.1 网络检索工具的分类	47
3.2.2 搜索引擎的主要类型	49
3.3 通用搜索引擎简介	49
3.3.1 国外通用搜索引擎	50
3.3.2 国内通用搜索引擎	66
3.4 垂直搜索引擎简介	76
3.4.1 垂直搜索引擎概述	76
3.4.2 垂直搜索引擎的关键技术	76
3.4.3 垂直搜索引擎的特征	77
3.4.4 垂直搜索引擎常用工具介绍	78
3.5 元搜索引擎简介	89
3.5.1 元搜索引擎概述	89
3.5.2 元搜索引擎的技术原理:资源整合/工具集成	91
3.5.3 元搜索引擎与普通搜索引擎的区别	91
3.5.4 元搜索引擎的优缺点	92
3.5.5 常用的元搜索引擎	93
思考题	103

第4章 互联网检索工具(二)——目录型网络检索工具	104
4.1 目录型网络检索工具概述	104
4.1.1 目录型网络检索工具的发展	104
4.1.2 目录型网络检索工具的特点	105
4.1.3 目录型网络检索工具的分类法与类型	106
4.1.4 目录型网络检索工具的结构和使用方法	107
4.1.5 目录型网络检索工具与搜索引擎的比较	109
4.2 目录型网络检索工具介绍	111
4.2.1 网络资源目录	111
4.2.2 学科信息门户	120
4.2.3 搜索工具门户	125
思考题	125
第5章 多媒体信息检索	127
5.1 图像的检索	127
5.1.1 图像检索的原理与技术	127
5.1.2 基于文本的图像搜索引擎简介	130
5.1.3 基于内容的图像搜索引擎	138
5.2 音频文件的检索	144
5.2.1 音频文件检索的原理与技术	144
5.2.2 基于文本的音频搜索引擎	147
5.2.3 基于内容的音频搜索引擎	151
5.3 视频文件的检索	154
5.3.1 视频文件检索原理	154
5.3.2 基于文本的视频搜索引擎	156
5.3.3 基于内容的视频检索	160
思考题	163
第6章 学术信息的网上检索与获取	164
6.1 图书信息的网上检索	164
6.1.1 世界各地图书馆的馆藏目录系统	164
6.1.2 联合目录数据库	166
6.1.3 网上书店	169
6.1.4 电子图书数据库和服务提供站点	172

6.2	学术期刊的网上检索	178
6.2.1	期刊出版信息的网上检索	178
6.2.2	期刊收藏信息的检索	180
6.2.3	期刊内容信息的检索	181
6.2.4	开放获取期刊	185
6.3	会议信息及会议文献的网上检索	189
6.3.1	会议信息检索	189
6.3.2	会议文献的检索	190
6.4	学位论文的网上检索	194
6.4.1	学位论文数据库检索	194
6.4.2	学位论文数字化项目资源	196
6.5	专利的网上检索	199
6.5.1	国外专利信息的网上检索	200
6.5.2	中国专利信息的网上检索	203
6.6	标准信息的网上检索	204
6.6.1	国外标准信息的网上检索	204
6.6.2	中国标准信息的网上检索	206
6.7	国际组织、政府机构及出版物信息的网上检索	209
6.7.1	国际组织机构信息的网上检索	209
6.7.2	国外政府信息的网上检索	211
6.7.3	中国政府信息的网上检索	215
	思考题	219
第7章	参考信息的网上检索	220
7.1	参考信息概述	220
7.1.1	参考信息检索概述	220
7.1.2	网上参考信息源的主要类型	220
7.2	百科知识的网上查询	221
7.2.1	百科知识查询概述	221
7.2.2	国内外网络百科全书简介	222
7.3	人物信息的网上查询	226
7.3.1	人物信息查询概述	226
7.3.2	网络传记资料简介	227
7.3.3	白页信息查询	228
7.4	地理信息的网上查询	232

7.4.1	地理信息查询概述	232
7.4.2	地名检索网站	232
7.4.3	地图的网上检索	233
7.4.4	旅游信息检索	237
7.5	时事、新闻的网上查询	242
7.5.1	时事、新闻信息查询概述	242
7.5.2	时事、新闻查询	243
7.5.3	网络版年鉴简介	246
7.6	机构信息的网上查询	248
7.6.1	机构信息查询概述	248
7.6.2	学术机构、政府机构名录	249
7.6.3	黄页信息服务	250
7.6.4	商务商机网站	255
7.7	语词信息的网上查询	257
7.7.1	语词信息查询概述	257
7.7.2	语文词典或语词查询网站选介	257
7.8	统计信息的网上查询	260
7.8.1	统计信息查询概述	260
7.8.2	统计信息检索网站选介	260
	思考题	265
第8章	互联网信息检索策略、技巧与提高	266
8.1	分析检索课题,明确信息需求	266
8.1.1	信息需求的描述	267
8.1.2	关键词的选取	268
8.1.3	关键词的组配	271
8.2	选择检索工具,实施检索策略	278
8.2.1	判断目标信息可能存在的地方	278
8.2.2	了解互联网信息检索的特殊性	279
8.2.3	了解可用的检索工具	279
8.2.4	仔细分析自己的信息需求并选择合适的工具	281
8.2.5	根据检索结果调整检索策略	285
8.3	进阶检索技巧	285
8.3.1	猜测 URL	285
8.3.2	右截断网址	286

8.3.3	利用网页快照·····	286
8.3.4	注意多义词·····	286
8.3.5	避免拼写错误·····	286
8.3.6	利用浏览器的“查找”功能·····	287
8.3.7	利用检索工具的特殊功能·····	287
8.3.8	使用辅助关键词·····	287
8.3.9	顺藤摸瓜·····	288
8.3.10	使用自然语言检索·····	288
8.3.11	注意单词的大小写·····	289
8.3.12	垂直检索·····	289
8.3.13	字段检索·····	289
8.3.14	善于利用错误信息·····	290
8.4	几个检索案例·····	291
8.4.1	年代久远的历史文件检索·····	291
8.4.2	寻找 KTV·····	293
8.4.3	煎饼果子翻译成英文·····	294
8.4.4	特定研究课题的资料收集·····	295
	思考题·····	297
第9章	网络信息检索的未来发展趋势·····	298
9.1	网络信息检索面临的困难与挑战·····	298
9.1.1	不断提高的用户需求·····	298
9.1.2	搜索引擎的“瓶颈”·····	300
9.1.3	网络信息检索中的法律风险·····	305
9.2	网络信息检索的发展展望·····	309
9.2.1	网络信息检索技术·····	310
9.2.2	网络信息检索服务·····	317
	思考题·····	322
	参考文献·····	323

CHAPTER 1

第 1 章

互联网信息资源

1.1 互联网上的信息资源

互联网已发展成为当今世界最大的信息平台,它是全球范围内发布、传播和交流海量科研信息、教育信息、商业和社会信息,实现信息资源共享的最主要的渠道,也是最大的信息宝库。人们要想在这浩瀚无边、变化多端而又鱼龙混杂的信息海洋中发现并查找出有利用价值的信息并不是一件易事。要完成确实、有效的网络信息检索,首先就必须对互联网上信息资源的分布、种类和利用价值等有较全面的认识和把握。

随着互联网普及率的提高,互联网正渗透到社会生活的方方面面,成为推动社会经济发展的重要引擎。在互联网发展初期,人们存在的对网络信息资源的误解和偏见虽已逐渐消退,但是仍然有必要建立对互联网信息资源全面、客观的认识。如:要了解互联网信息资源并非应有尽有、唾手可得;要意识到在互联网上找到的、下载的信息并非都是正确的;要明确互联网在整个信息生态环境、信息交流链中所处的位置,了解网络信息资源的特点和利用价值;同时还应掌握各类网络搜索工具的特点、功能和查找方法;进而在面对各种媒体、各种信息系统、搜索工具时,能够根据检索需求的不同,做出正确的选择。

在当今的信息环境下,能够“在适宜的时机,以适宜的形式、适宜的价格找到确切的信息”是具有较高检索水平的重要体现。

1.1.1 网络上的信息内容

作为检索、利用互联网信息资源综合策略的第一步,就要了解网上主要有哪些信息内容,哪些信息内容是不容易被发现和查找到的。互联网上主要的信息内容包括以下几种。

1. 政府信息

政府信息是指国际组织、各国政府及其相关部门所发布的信息。包括:有关组织机构的宗旨、业务范围、人员、出版物、最新消息发布;还有各种法律、法规或相关政策信息;从统计数据到立法建议等,内容相当广泛。政府信息可以说是网上最有价值的一部分信息,由于出自官方,所以具有权威性、可靠性等特点。另一个突出的特点是价廉,其中大部分信息都是免费的,网民可以免费获得许多没有版权限制的政府文件。

2. 科研信息

科研信息主要指各专业学术机构(如学会、协会、研究院所等)所设立的网站及其相关信息,包括目标、宗旨、成员、主要出版物、最新学术活动安排(如会议、展览等)、各种学术性电子期刊、学科专题论坛、讨论组,各学科应用的工具性计算机软件,还有近年来兴起的许多开放存取资源等。

3. 教育信息

教育信息主要有各大学所设立的网站及相关信息。其内容相当广泛,有该校各学院、系、专业的介绍,学位、奖学金的设立,入学申请表、校历,以及学校周边环境、生活设施、公共交通。还有各学科专业的教学计划、课程表,以及教师的个人网页(通常含教师个人简历、所任课程的教学大纲、研究成果、著述、近期研究活动等)。另外,互联网已经发展成为一种重要的教育手段、学习环境,它所包含的教育信息是极为丰富的,如:各种远程教育的课件;用于学生自我学习的交互式虚拟教室;还有为教师教学而提供的大量参考资料与素材,如有关科学、艺术和人类感知的各种图片、照片等多媒体网站及数据库。

4. 媒体信息

媒体信息主要包括各类信息媒体(如报纸、杂志、电视、广播等)的网站资源,世界各地图书馆的数字化馆藏及其他提供各类电子书、数字期刊等信息内容的网站。近年来,互联网作为新型媒体快速崛起,网络媒体受众广泛、关注度高、交互性强、价格低廉,在满足人们不同信息消费需求方面具有独特优势,已成为最具活力、最具发展潜力的大众媒体。

5. 商务信息

目前,互联网正从信息传播和娱乐消费为主向商务服务领域延伸,电子商务的迅速发展使得互联网开始逐步深入到国民经济的更深层次和更宽领域,网络上各种与商务活动有关的信息也日渐增多,既有各种原材料、制成品的供求信息、价格行情,也有各种市场调查、市场研究与分析预测信息,还有各种采购、招标、推广与服务信息等。随着电子商务在整体网络经济中的占比继续扩大,网络商务信息也将不断丰富和繁盛。

6. 休闲娱乐信息

休闲娱乐信息曾被认为是网上“最成功”并具有占比优势的领域,包括各种网络游戏、网络文学、音乐、视频信息。有为任何一种可想象出的兴趣、爱好(如足球、音乐、电影、收藏、烹饪等)而开设的各种网站、讨论组、俱乐部等,也有各类旅游信息,风景名胜、特产、民俗,以及宾馆、饭店、住宿、交通等信息。

1.1.2 网上不容易获取的信息内容

尽管网上信息发展繁盛,事实上还是有许多重要信息不能在网公开,或不能让用户自由访问、获取,比如有关企业的战略,发展规划,技术、商业、贸易秘密等。各类有知识产权限制的文学、艺术作品、技术成果和数据库等,一般都要经过付费、订阅等途径获得认证、授权后,才能访问、浏览或下载到相关内容。

另外,应了解到互联网上还存在着“看不见的网页”(The Invisible Web,或称隐形网页),简单来说就是指由于各种原因,不能被通用搜索引擎(如 Google、百度等)搜索到的网页。这部分网络资源曾被认为有很大数量。Bright Planet 公司曾估计,相比较于被普通搜索引擎收录的 10 亿网页,仍有 5 500 亿网页未被收录^①。该公司近期网页数据显示:据早前的估计,这部分隐身网页的数量是被收录网页数量的 4 000~5 000 倍。当然,随着搜索引擎技术的不断发展,以及搜索服务商与内容提供商之间合作的加强,许多过去隐身的网页现在开放了、可以被检索和使用了。但网络用户还是应意识到:并非所有的网络资源都能够通过搜索引擎检索到,而搜索不到的网页或信息内容并非在网络上不存在。

1.2 互联网信息资源的特点及局限性

互联网作为数字化、网络化信息的核心和集成,它与传统的信息媒体和信息交流渠道相比有很大的不同。用户在享用互联网信息资源的同时,要了解其特点、优势和局限性,才能使互联网信息资源的价值能够得到更好的发挥。

1.2.1 互联网信息资源的特点

互联网信息资源的特点可以概括为以下几点。

- 广泛性。互联网连接了世界范围数以亿计的计算机,信息容量巨大,传播广泛。互联网上的信息资源极为丰富,覆盖面广,涵盖了各种学科、领域,几乎无所不包。

- 多样性。互联网以超文本、超媒体、集成式地提供多种信息类型,除文字信息外,还有声音、图像等音频、视频信息,人们可通过网络搜索到歌曲、文章、电影等多种形式的信息资源。

- 共享性。互联网是一种共享程度极高的信息资源配置平台,由于其信息存储形式及数据结构具有通用性、开放性和标准化的特点,使得网络信息资源在时间和空间上得到了最大程度的延伸和扩展。一份信息资源上网后,不仅可以及时地提供给本地用户,而且可以通过网络传播到世界各地供所有网络用户共享。

- 新颖性。互联网提供了辐射全世界的高速信息传输通道,网络信息具有出版周期短,时效性强的特点。网站、网页会随着时间的推移而不断更新,同时许多新的互联网交流形式(如推特/微博、博客等)更丰富了信息的传播渠道,加快了信息的传播速度,对一些突发事件、社会新闻、科学发现报道的原创率、首发率甚至超越了传统的新闻媒体和出版发行系统,已经成为追求信息新颖性、时效性的用户不可忽视的信息平台。

- 互动性。网络信息资源具有交互性、互动性,用户既是网络信息资源的利用者,也是网络信息资源的开发主体和建设者。以网络为媒介,世界各地的学者可以通过加入网

^① “看不见的网页”介绍, <http://www.sowang.com/SOUSUO/20041130.htm>

络社区、群组交流探讨研究心得；教师和学生则可以利用在线学习平台交流学习资料，营造开放、自由的学习气氛；用户可以通过在线问答系统获得其他用户的帮助，还可以利用自己所掌握的知识为他人释疑解惑。

- 经济性。互联网是一种比较经济、廉价的信息获取方式，用户获取网络信息资源大多是免费的，一般只需付出查找信息的时间成本、一定的通信费用和打印费用。对比传统的通过订阅、购买印刷型出版物获取信息的方式，在时间和费用上都是比较经济的。

1.2.2 互联网信息资源的局限性

网络用户在享受互联网信息平台的方便、快捷时，也要意识到互联网信息资源的局限性。要注意选择相关的信息工具、掌握检索网络信息的方法，有效地查找、获取信息，并对所获取信息资源的信息价值做出客观判断，才能充分地利用网络信息。互联网信息的局限性主要体现在以下几方面。

- 分散。网络信息资源来自世界各地数以亿计的计算机，信息生产及传播渠道的分散性很强。既有组织机构、企业发布的信息，也有用户的原创；互联网在为使用者提供了广阔信息空间和丰富信息机会的同时，也使用户难以把握所需信息的来源和传播渠道。

- 无序。由于大量信息资源分散在网络的不同层次和节点，不具有中心点，没有统一的管理机构和发布标准，致使信息资源缺乏组织、整理，处于无序状态。

- 多变。网络信息资源传播迅速，动态性、变化性很强，大部分资源站点更新非常频繁，网页的变化、更迭、新生、消亡等随时随地都在发生，致使某些信息内容稍纵即逝、难以捕捉。

- 信息质量难以控制。由于信息创造者的水平处于不同层次，致使生产出的网络信息资源形式多样，质量良莠不齐。

- 信息安全难以保障。伴随着电子商务市场的扩大和交易额的增长，网络欺诈活动日益猖獗，用户在利用网路资源时常常会遭遇到虚假网站、身份窃取、账户丢失、黑客攻击等事件，使个人隐私及财产受到侵犯和威胁。

1.3 互联网上信息资源的种类

互联网信息资源丰富多彩、包罗万象、分布广泛，呈现着无限、无序的特点。了解互联网信息资源的大致分类，有助于用户理清网络信息资源的分布、交流渠道，认识不同类型、不同信息价值网络资源的传播、获取方式等。对互联网信息资源的分类可以从多个不同的角度进行，例如，根据媒体类型的不同，可分为文本、图形、图片、声音、音乐、电影及各种视频影像等多种媒体格式的信息；根据使用权限的不同，可分为开放信息（无访问、使用限制，可自由访问的网络信息）与限制访问信息（即有一定访问限制，需要付费订阅或注册登记获得授权以及账户密码后才能访问、使用的网络信息）；根据信息的表现形式，可以分

为全文型(如各种报纸、杂志)、数据型(如各种报价网站)、目录型(如图书馆藏书目录)、事实型(如个人主页)等。而较常见的还是按照信息资源所采用的网络传输协议的不同来划分。

1.3.1 Web 信息资源

Web 是 World Wide Web(WWW,3W)的缩写,中文译为“全球信息网”、“万维网”。它起源于 1989 年 3 月欧洲量子物理实验室 CERN(European Laboratory for Particle Physics)的科学家 Tim Berners Lee 提出的一份立项报告,建议采用超文本技术设计分布式信息系统并把 CERN 内部的各个实验室连接起来,在系统建成后,将可能扩展到全世界。同年夏天, Tim Berners Lee 成功开发出世界上第一个 Web 服务器和第一个 Web 客户机,虽然还非常简陋,但它确实是一个所见即所得的超文本浏览/编辑器;1989 年年底,该发明被正式定名为 World Wide Web。由于用户在通过 Web 浏览器访问信息资源的过程中,无须再关心一些技术细节,而且界面非常友好,人们只要通过输入、点击、链接等简单方法,就可以很便捷地浏览、获得丰富的信息资料。因而 Web 在互联网上一经推出就受到热烈欢迎,引起广泛关注和大量推广使用。1994 年互联网上传送的 WWW 数据量首次超过 FTP 数据量,成为访问网络资源的最流行的方法。至今,Web 服务器已成为互联网上最大的计算机群,Web 文档之多、链接的范围之广,令人难以想象。可以说,WWW 技术给互联网注入了强大的发展动力,使互联网的普及迈出了开创性的一步。

WWW 之所以受到人们的欢迎,在于其高度的集成性。它把各种类型的信息(如文本、声音、动画、录像等)和服务(如 News、FTP、Telnet、Gopher、Mail 等)无缝链接,提供了丰富多彩的图形界面,直观方便;其信息服务支持超文本和超媒体;可使用户在全世界范围内查询、浏览最新信息。Web 信息由资源地址域名和 Web 网站、站点组成,Web 站点可以相互链接,以提供信息查找和漫游访问,还可以使用户与信息发布者或其他用户相互交流信息。掌握 Web 资源的使用需要了解以下几个重要概念。

1. 超文本传输协议(HTTP)

HTTP 是浏览器与 Web 服务器之间相互通信、传输、响应用户请求的协议。Web 服务器根据客户提出的 HTTP 请求,为用户提供信息浏览、数据查询、安全验证等方面的服务。

2. 超文本标记语言(HTML)

HTML 是一种专门编程语言,具体规定和描述了文件显示的具体格式。也即 Web 的描述语言,用于编制通过 WWW 显示的超文本文件。

3. 通用资源定位程序(URL)

通用资源定位程序又称网络资源的统一定位格式或统一资源定位器。实际上就是一个用以标识文档类型及其所在网络地址的字符串,它的用途是用统一的方式指明互联网上信息资源的位置。URL 包括 3 部分:所使用的传输协议、服务器名称、该服务器上定位文档的全路径名。URL 不仅用于 HTTP 协议,还适用于 FTP、Gopher、Telnet 等协议。

4. 主页(Home Page)

主页是浏览器访问某个 Web 服务器上的信息时第一个链接到的文档。它是该服务器入口处的 HTML 文件,是进入该服务器、访问其所提供的各类信息资源的导引页,要为用户浏览该服务器中的有关信息提供方便。

5. 链接(Link)

链接是 Web 网页的元素,是指向其他信息资源的指针,把 Web 页捆绑在一起。跟随链接可从一个文档跳到另一个相关文档。

6. 浏览器

浏览器指 Web 服务客户端的浏览程序,是显示网页服务器或档案系统内的文件,并让用户与这些文件互动的一种软件。可向 Web 服务器发送各种请求,并对从服务器发来的超文本信息和各种多媒体数据格式进行解释、显示和播放。个人计算机上常见的网页浏览器包括微软的 Internet Explorer(IE)、Mozilla 的 Firefox、Google 的 Chrome、苹果的 Safari 等。国内也有搜狗浏览器、傲游浏览器、百度浏览器等多家。大部分浏览器支持除了 HTML 之外的广泛格式,如 JPEG、GIF 等图像格式,并且能够扩展支持众多的插件(plugin)。另外,许多浏览器还支持其他的 URL 类型及其相应的协议,如 FTP、Gopher、HTTPS(HTTP 协议的加密版本)。HTTP 内容类型和 URL 协议规范允许网页设计者在网页中嵌入图像、动画、视频、声音、流媒体等。

目前,WWW 是互联网上最为流行的信息传播方式,Web 资源成为互联网上最具优势的主流信息资源。为满足各种不同的信息发布和交流要求,Web 资源呈现了快速、多样化的发展,有各种不同的网站形式,如门户网站、社交网站、搜索网站、商务网站、教育网站、新闻网站、内容网站、数据库网站等;还有适合不同信息传播需要的网站功能,如博客(Blog)、微博(twitter,推特)、聚合内容(RSS-Really Simple Syndication)等。

1.3.2 Telnet 资源

Telnet 协议是互联网远程登录服务(Remote Login)的标准协议和主要方式,其功能是通过在远程计算机上登录,使用户正在使用的计算机暂时成为它所登录的某一远程主机的仿真远程终端,进而实时访问、共享、使用到远程计算机系统中对用户开放的相应资源。可以通过 Telnet 方式使用的资源既包括硬件资源,如超级计算机、精密绘图仪、高速打印机、高档多媒体输入/输出设备等;也包括软件资源,如大型的计算程序、图形处理程序,以及大型数据库、电子图书与档案等内容信息资源。虽然这些资源一般都存储在远程或异地的计算机系统中,但通过互联网使用起来却像在身边一样方便。

使用 Telnet 协议进行远程登录时需要满足以下条件:在本地计算机上必须装有包含 Telnet 协议的客户端程序;必须知道远程主机的 IP 地址或域名;必须有登录标识与口令。访问 Telnet 信息资源的一般过程是:

(1) 在 Telnet 命令中输入远程计算机的域名或 IP 地址,在本地机与远程主机间建立